DOCUMENT RESUME

ED 304 451 TM 012 682

AUTHOR Love, Gloria C.
TITLE Understanding Experimentwise Error Probability.
PUB DATE Nov 88
NOTE 19p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Louisville, KY, November 9-11, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Analysis; *Error Patterns; Estimation (Mathematics); Experiments; *Hypothesis Testing; *Probability; *Statistical Analysis
IDENTIFIERS Error Analysis (Statistics); *Experimentwise Error; *Type I Errors

ABSTRACT

The probability of experiment-wise error is explored. Overall, the experiment-wise error rate is directly related to the test-wise error rate--the alpha level set by researchers to curtail the existence of a Type I error. A Type I error occurs when a true null hypothesis is rejected in a given study or experiment. The experiment-wise error rate is a result of the increased possibility for Type I error; it emerges when several hypotheses are tested at the same alpha level within a single study. The experiment-wise error rate may be substantially above the alpha level used to test each hypothesis separately. A single sample chi square test, using coin tossing as an example, illustrates the implications of the inflated experiment-wise error rate. Two tables present the experimental data and an attachment includes a program for calculating experiment-wise error inflation rates. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED304451

UNDERSTANDING EXPERIMENTWISE ERROR PROBABILITY

Gloria C. Love

University of New Orleans

and

Orleans Parish Schools

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gloria C. Love

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The author acknowledges the helpful assistance of Theresa DuRapau, RSM in preparing the derived proof of the formula for experimentwise Type I error rate. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville, Kentucky, November 10, 1988.

TM012682

ABSTRACT

Researchers set alpha levels in order to curtail the existence of Type I errors. The alpha level set by researchers for this purpose is the testwise alpha rate. The testing of several hypotheses in a single study inflates the possibility of a Type I error occurring somewhere in the study or experiment. The experimentwise error rate is the result of the inflation. The experimentwise error rate may be substantially above the alpha level used to test each hypothesis separately. A single sample chi square test was used to illustrate the implications of the inflated experimentwise error rate.

UNDERSTANDING EXPERIMENTWISE ERROR PROBABILITY

The testwise error rate is the alpha level used in testing a given null hypothesis. The experimentwise error rate emerges when several hypotheses are tested at the same alpha level within a single study. Sometimes the two rates are the same and sometimes the rates are not. The experimentwise error rate is calculated dependent on the value of the testwise error rate as well as the number of hypotheses in a given study.

The testwise probability of making errors is controlled by choosing an appropriate alpha level of statistical significance. The alpha level normally chosen by researchers is $p = .05$. If the alpha level is set at the .05 level of statistical significance then five percent of the time the null hypothesis will be rejected. The null hypothesis will be rejected five percent of the time even though no differences exist in the population means.

A Type I error involves the rejection of a true null hypothesis in a given study. The null hypothesis is rejected based on the alpha level set by the researcher. Researchers are aware that Type I errors are generally undesirable, and typically set alpha probabilities to be

very low to make Type I errors very unlikely (Fehr, 1964; Fish, 1988; Ingram, 1977). The possibility of a Type I error exists each time a hypothesis test is performed. Kerlinger (1979) gives an example:

> Suppose we conduct a large number of experiments and use the same alpha level, p = .05, for each experiment. In any one of these experiments, the chances are 1 in 20 that we will make a Type I error when the null hypothesis is true. Another way of looking at the Type I error is to consider the number of Type I errors we may expect on the average when the null hypothesis is true. This number is found by multiplying alpha times the number of experiments; if we conduct 20 experiments then we expect to make one Type I error, .05(20), and if we conduct 100 experiments, then we expect to make five Type I errors, .05(100). It is obvious that by increasing the number of experiments we conduct and analyze, we are increasing the number of Type I errors as well. The more hypotheses included in a given study the

greater is the risk for multiple Type I errors. Experimentwise Type I errors involve the probability of making an error somewhere in the study. Researchers differentiate between what is called the test-wise alpha level and the experiment-wise alpha level (Gravetter & Wallnau, 1985).

In an experiment, the testwise error rate may be the same as the experimentwise error rate, provided there is only one hypothesis being tested or if dependent variables are perfectly correlated (Fish, 1988). If, in an experiment, several comparisons are made, each at alpha = .05, the testwise rate would be used for each hypothesis. The probability of making a Type I error would be .05 for each separate comparison. However, considering the experimentwise error rate in a multiple set of comparisons, there exist a greater probability of making several Type I errors in the overall set of comparisons. It is well known that performing several univariate tests when a study involves multiple dependent variables inflates the probability that a Type I error will occur somewhere in the study (Thompson, 1988).

The relationship between the two error rates is expressed by

$$\text{alpha(ex)} = 1 - (1 - \text{alpha})^{c}$$

where c represents the number of independent comparisons that are conducted (Gravetter & Wallnau, 1985; Kerlinger, 1986; Thompson, 1988). The experimentwise alpha(ex) is dependent on the testwise alpha value set by the researcher. The hypotheses must remain at the same testwise alpha level within a single study. See Table 1 for values of testwise alpha and the inflated experimentwise alpha at different levels of testwise alpha. The table illustrates the alpha values for a one sample research study. As the value of the testwise alpha decreases, the value of the experimentwise error rate decreases. As the number of tests increase, the value of the experimentwise error rate increases. Whenever the dependent variables in a study are perfectly uncorrelated the experimentwise error rate is inflated to the maximum. The experimentwise error rate formula may be derived algebraically.

The algebraic derivation of the error rates relationship is as follows:

Using the probability rule for multiplication,

$$P(A \text{ and } B) = P(A) \times P(B)$$

$$P(\text{Type I error(s)}) = 1-(1-\text{alpha})^{c}$$

where c is the number of independent comparisons. Using the smallest comparison, assign 2 to c, set the alpha level at p = .05 and two null hypotheses H1 and H2. -H1 and -H2

represent the rejection of the hypotheses when true. H1 and H2 are not of concern since the rejection of at least one of the hypotheses is required. The probability now is

P(Type I error(s))

= (-H1 and H2 or H1 and -H2 or -H1 and -H2)

=P(-H1 and H2) + P(H1 and -H2) + P(-H1 and -H2)

=P(-H1) X P(H2) + P(H1) X P(-H2) + P(-H1) X P(-H2)

= (.05) X (.95) + (.95) X (.05) + (.05) X (.05)

= .05(.95 + .95 + .05) Distributive Property

= .05(.95 + 1) Substitution Property

= .05(.95) + .05 Distributive Property

= .95(.05) + .05 Commutative Property

= .95(1-.95) + .05 Substitution Property

= .95-(.95)(.95) + .05 Distributive Property

$= .95 + .05 - (.95)^2$ Commutative Property & Substitution Property

$= 1 - (1 - .05)^2$ Substitution Property

Therefore,

$P(\text{Type I error(s)}) = \text{alpha(ex)} = 1-(1-\text{alpha})^c$.

This formula only holds for independent comparisons.

Consider the following examples. If the testwise error rate = .05 and c = 4 comparisons then

$$\text{alpha(ex)} = 1-(1-\text{alpha})^{c} = 1-(1-.05)^{4}$$

$$= 1-.8145 = .1855.$$

Insert Table 1 about here

The probability that a Type I error will occur somewhere in the four tests of hypotheses is between 0.05 and 0.19.

The experimentwise error rate is approximated by

$$\text{alpha(ex)} \sim c(\text{alpha})$$

This is an approximation and is only accurate to some degree for small values of alpha and for small numbers of comparisons. In the given example,

$$\text{alpha(ex)} \sim c(\text{alpha})$$

$$\sim 4(.05) \sim .2 \text{ is compared to } .1855 \text{ or } .19.$$

Ironically, experimentwise error rate is at a maximum when the hypotheses tested within an experiment are orthogonal or uncorrelated (Thompson, 1987). Thompson provides an example:

The tests of all omnibus hypotheses in

a factorial multi-way ANOVA with equal numbers of subjects in each cell are all uncorrelated. This is the reason why the sums of squares (SOS) for each effect plus the error SOS add up to exactly equal to the SOS total.

Null hypotheses may be tested in many ways. Some tests are the t-test, the F test, and the chi-square test. The present paper discusses the chi-square test and the experimentwise error rate. The chi-square test is a statistical test used to evaluate hypotheses of association for contingency table data. Two sets of frequencies are compared in the chi-squared test. Frequencies resulting from direct observation are "observed frequencies" and those used for comparison are "expected frequencies". Expected frequencies are considered theoretical frequencies.

A binomial experiment is one of the simplest experiments since each outcome can take only one of two possible values. The basis for binomial experiments is the Bernoulli trial. This requires that the outcome of a single trial is either a success or a nonsuccess and that the chance of success is the same for every trial (Ingram, 1977). Researchers are generally concerned with many Bernoulli trials when tossing a fair coin.

Using the null hypothesis that a coin being tossed 100

times will yield the same percentage of heads and tails, the probability of getting a head is 50% and the probability of getting a tail is 50%. Therefore P(H)=1/2 and P(T)=1/2. The null hypothesis will be rejected at the .05 testwise alpha level provided chi square > 3.84. Because there are two classes which require two restrictions, degrees of freedom (df) = k - 1 = 2 - 1 = 1.

The expected frequencies are:

Heads	50	100
Tails	50	

Suppose the observed frequencies are:

Heads	40	100
Tails	60	

To determine whether the differences between the expected and observed frequencies are statistically significant, apply the chi square formula for the test of independence. The formula is the same for the goodness of fit test. The difference between the observed and expected frequencies are computed, squared for each cell and divided by the expected frequency for that cell.

$$X^2 = \sum_{i=1}^{k} \left[\frac{(fo - fe)^2}{fe} \right]$$

where X^2 = chi square, fo = observed frequency in class(i), fe = expected frequency in class(i) and k = the number of classes.

$$X^2 = \sum_{i=1}^{2} \left[\frac{(fo - fe)^2}{fe} \right] = \frac{(40 - 50)^2}{50} + \frac{(60 - 50)^2}{50} = 4.00$$

The null hypothesis is rejected at the .05 testwise alpha level since chi square = 4.00 and 4.00 > 3.84. The data indicates a significant difference in the observed number of heads and the expected number at the testwise alpha level of .05. If the 100 tosses were repeated 10 times with the same testwise alpha level then

$$\text{alpha(ex)} = 1-(1-\text{alpha})^{10} = 1-(1-.05)^{10} = .4013 \text{ or } .40.$$

The computer was used to illustrate other values of chi

square for fifty 100 random tosses of a coin, as indicated in Table 2. For 50 tests performed at a .05 testwise error rate, the experimentwise error rate may be as high as .9230 or .92, as noted in Table 1. The researcher is 92% sure a Type I error occurred somewhere in the 50 tests. As noted in Table 2, one Type I error actually did occur within the set of 50 tests, as expected.

Insert Table 2 about here

SUMMARY

Overall, the experimentwise error rate is directly related to the testwise error rate. The testwise error rate is the alpha level set by researchers in order to curtail the existence of a Type I error. A Type I error occurs when a true null hypothesis is rejected in a given study or experiment. The alpha level used in testing a given hypothesis is called the testwise alpha. The testing of several hypotheses in a single study inflates the possibility of making a Type I error somewhere in the study. The experimentwise error rate is a result of the inflation. The experimentwise error rate is substantially above the alpha level used to test each hypothesis separately. The chi square test was utilized to illustrate the inflation of

the experimentwise error rate.

REFERENCES

Fish, L. J. (1988). Methods, plainly speaking: Why multivariate methods are usually vital. Measurement and evaluation in Counseling and development . vol 21,3.

Fehr, H. F., Blunt L. N. , & Grossman, G. (1964). An introduction to sets, probability and hypothesis testing. Boston: Heath.

Gravetter, F. J. & Wallnau, L. B. (1985). Statistics for the behavioral sciences. New York: West.

Ingram, J. A. (1977). Elementary statistics . Massachusetts: Benjamin/Cummings.

Kerlinger, F. N. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart and Winston.

Kerlinger, F. N. (1986). Foundation of Behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Thompson, B. (1987). The importance of a priori contrasts in analysis of variance research. Paper presented at the annual meeting of the Mid-South Educational Research Association. Mobile.

Thompson, B. (1988). Misuse of chi-square contingency table test statistics. Educational and Psychological Research. Vol 8,1.

Table 1

EXPERIMENTWISE ERROR INFLATION RATES

TESTWISE ALPHA LEVELS			
N	.10	.05	.01
1	.1	.0499999996	9.9999995E-03
2	.19	.0974999993	.0198999991
3	.271	.142624999	.0297009987
4	.3439	.185493749	.0394039883
5	.40951	.226219061	.049009948
10	.65132156	.401263059	.0956179209
15	.794108868	.536708767	.13994164
20	.878423346	.641514075	.182093055
25	.928210201	.722610425	.222178632
30	.957608842	.785361234	.260299617
35	.974968445	.833916614	.296552294
40	.985219117	.871487842	.33102823
45	.991272036	.900559742	.363814501
50	.994846225	.923055023	.394993919
55	.996956747	.940461444	.424645236
60	.99820299	.9539302	.452843343
65	.998938883	.964352067	.479659462
70	.999373421	.972416309	.505161325
75	.999630012	.978656266	.529413343
80	.999781525	.983484625	.55247677
85	.999870993	.987220718	.574409861
90	.999923823	.990111635	.595268011
95	.999955018	.992348572	.615103905
100	.999973439	.994079471	.633967642

N=Number of tests
Ex=1-(1-Testwise Alpha)^N

Table 2

CALCULATIONS OF CHI SQUARE FOR THE COIN TOSS

N	HEADS	TAILS	CHI SQUARE
1	47	53	.36
2	52	48	.16
3	43	57	1.96
4	49	51	.04
5	52	48	.16
6	59	41	3.24
7	52	48	.16
8	50	50	0
9	53	47	.36
10	42	58	2.56
11	39	61	4.84 ←
12	56	44	1.44
13	46	54	.64
14	58	42	2.56
15	54	46	.64
16	50	50	0
17	56	44	1.44
18	49	51	.04
19	51	49	.04
20	52	48	.16
21	48	52	.16
22	45	55	1
23	51	49	.04
24	41	59	3.24
25	50	50	0
26	56	44	1.44
27	51	49	.04
28	53	47	.36
29	46	54	.64
30	43	57	1.96
31	53	47	.36
32	49	51	.04
33	56	44	1.44
34	48	52	.16
35	49	51	.04
36	44	56	1.44
37	45	55	1
38	52	48	.16
39	50	50	0
40	49	51	.04
41	50	50	0
42	50	50	0
43	48	52	.16
44	45	55	1
45	51	49	.04
46	50	50	0
47	47	53	.36
48	53	47	.36
49	55	45	1
50	48	52	.16

N=Number of tests. Each includes 100 tosses.
The sum of the overall chi squares= 37.44

```
10  REM   PROGRAM FOR CALCULATING EXPERIMENTWISE ERROR INFLATION RATES
20  PRINT  TAB( 20);"Table 1"
30  PRINT
40  PRINT "EXPERIMENTWISE ERROR INFLATION RATES"
50  FOR I = 1 TO 46: PRINT "_";: NEXT I
60  PRINT "TESTWISE ALPHA LEVELS"
70  PRINT "N"; TAB( 09);".10"; TAB( 24);".05"; TAB( 37);".01"
80  FOR I = 1 TO 46: PRINT "_";: NEXT I
90 A = .10:E = .05:D = .01
100  FOR I = 1 TO 4
110 E1 = 1 - (1 - A) ^ I
120 E2 = 1 - (1 - E) ^ I
130 E3 = 1 - (1 - D) ^ I
140  PRINT I;"   ";E1; TAB( 20);E2; TAB( 34);E3
150  NEXT I
160  FOR I = 5 TO 100 STEP 5
170 E1 = 1 - (1 - A) ^ I
180 E2 = 1 - (1 - E) ^ I
190 E3 = 1 - (1 - D) ^ I
200  PRINT I;"   ";E1; TAB( 20);E2; TAB( 34);E3
210  NEXT I
220  FOR I = 1 TO 46: PRINT "_";: NEXT I
230  PRINT
240  PRINT "N=Number of tests"
250  PRINT "Ex=1-(1-Testwise Alpha)^N"
260  END
```

```
10  REM   PROGRAM FOR RANDOMLY GENERATING CHI SQUARE VALUES
20  PRINT "                    Table 2"
30  PRINT
40  PRINT "CALCULATIONS OF CHI SQUARE FOR THE COIN TOSS"
50  FOR I = 1 TO 44: PRINT "_";: NEXT I
60  PRINT
70  PRINT "N"; TAB( 4)"HEADS","TAILS","CHI SQUARE"
80  FOR I = 1 TO 44: PRINT "_";: NEXT I
90  PRINT
100  FOR I = 1 TO 50
110 H = 0:T = 0:E = 50:X2 = 0
120  FOR J = 1 TO 100
130 X =  INT (2 *  RND (1))
140  IF X = 1 THEN 170
150 T = T + 1
160  GOTO 180
170 H = H + 1
180  NEXT J
190 X2 = ((H - E) ^ 2) / E + ((T - E) ^ 2) / E
200 X2 =  INT ((X2 + .005) * 100) / 100
210  PRINT I;"  ";H,T,X2
220 SX2 = SX2 + X2
230  NEXT I
240  FOR I = 1 TO 44: PRINT "_";: NEXT I
250 SX2 =  INT ((SX2 + .005) * 100) / 100
260  PRINT
270  PRINT "N=Number of tests. Each includes 100 tosses.
280  PRINT "The sum of the overall chi squares= ";SX2
290  END
```